Proceedings of the International Conference of the Digital Humanities Association of Southern Africa (DHASA) 2019

Pretoria, South Africa

25th - 29th of March 2019

Short Papers

In Search of Missing Narratives and the process of archiving gaps

Tinashe Mushakavanhu

Keywords: black archives, Zimbabwe, intellectual history, digital infrastructure

Abstract

In 2016 two young Zimbabwean educationists living and working in the United States embarked on building reading zimbabwe.com a webography mapping the intellectual history of Zimbabwe since 1956 when the rst black intellectuals published books. In the process they encountered many gaps – books out of print, books printed overseas, books missing in libraries. The major casualty was books on Zimbabwe written in the vernacular – Shona and Ndebele. They were missing wholesale on the internet, yet deposited at libraries at rich universities such as Yale or Harvard. This paper will discuss how deleterious the set up of the internet is for marginal cultures and languages using infographics and maps generated through readingzimbabwe.com.

Library Support for Humanities Research Data Management: creating a collaborative environment

Thozama Bici, Mark Snyders

Keywords: Digital humanities, Higher education, University libraries, Research Data Management, Research data

Abstract

Library support for Research Data Management requires a detailed understanding of data being collected in the Humanities research environment. The aim of this study is to identify how subject librarians could collaborate and foster greater ties with academics and researchers in the humanities by introducing and providing training for various support services and facilities. A survey was conducted to categorise types of data collected in the humanities and how it is stored and shared. The study addresses data management planning, storage and sharing practices. Ethical issues including ownership and sharing of collected data are also addressed. The results of the study reveal levels of understanding of research data management and o ers solutions to how subject librarians could support humanities researchers.

How to conduct a multi-scale digital humanities project: using computer vision, arti cial intelligence and digital visualisations to combine close and distant humanities research

Karli Brittz

Keywords: multi-scale, arti cial intelligence, computer vision, digital methodology

Abstract

New media theorist Lev Manovich argues that the key aspect of digital humanities is combining 'distant reading' of patterns with 'close-readings' of particular phenomenon, by means of a multi-scale reading. Through the demonstration of such a multi-scale digital humanities project, this talk aims to establish the importance of both distant and close readings of humanities research enabled

through digital technologies. In doing so, the talk will provide methodological guidelines to undergo such a project. Finally, the paper will also showcase the use of arti cial intelligence or machine learning software to analyse a large amount of images to identify visual patterns and a critical analysis of these patterns.

As part of a research project investigating the human-nonhuman relation between humans and their companion species as it manifests on social media, a digital project, Pupparazzi, was created that uses computer vision and machine learning to analyse a large data set of images from the social media platform Instagram. These images are then visualised on a digital platform to reveal patterns and critical details regarding selected images. To demonstrate the methodological approach and bene ts of such a multi-scale digital humanities project, the talk will use Pupparazzi as a speci c example. In addition the digital project will also be used to show the techniques and software employed in the creation of Pupparazzi, which includes:

(1) creating and extracting a dataset of images from Instagram based on random selection

(2) running this dataset through recognition and analytic software, which provide algorithmically calculated estimates of commonalities in the images (for example, position of dog, close-up images and top-down photographs)

(3) extracting metadata from the images in the dataset based in the social media platform regarding time, place and other formal elements

(4) visualising this metadata and data using big data visualisation computational tools

(5) making critical deductions or interpretations based on these visualisations and results

Finally, by sharing the custom-made software and tools as well as a basic strategy and lessons learnt from a multi-scale digital humanities project, the talk aims to add to the developments in the elds of digital humanities and image studies and share new tools for others to use within their respective elds.

The application of geographic information systems (GIS) to visualise the nineteenth-century life histories of South African people with intellectual disabilities

Rory du Plessis

Keywords: GIS, visualisation, historical data

Abstract

There is a disconcerting dearth of scholarship of the life histories of South African people with intellectual disabilities from the nineteenth century. I seek to correct the “historical invisibility” (Jackson 2013: 88) of people with intellectual disabilities by exploring the life stories of several people who were patients at the Institute for Imbecile Children (est. 1894), in Makhanda, South Africa, from 1894 to 1907. The casebooks for the Institute provide a valuable resource for studying the life stories of the patients as they provide a window into exploring the social worlds of

a patient, their course of life, and pathways to the Institute (Coleborne 2009: 71; 2010: 14). In the study, I employ geographic information systems (GIS) to visualise the pathways and life course of several patients (see also Travis 2010, 2013 and 2014). By nding "patterns, facilitat[ing] comparisons, enhanc[ing] perspective, and illustrat[ing] data", GIS o ers the study a "powerful tool in the … analysis of evidence" (Bodenhamer 2007: 107). The GIS visualisations identi ed the unique contours, complexity, and speci c sites presented in each patient's life course and foster a closer reading of the casebooks to understand and "explain the historical processes that form[ed] these patterns" (Gregory and Ell 2007: 118).

References

Bodenhamer, D.J., 2007, 'Creating a landscape of memory: the potential of humanities GIS', International Journal of Humanities and Arts Computing, 1(2), 97–110.

Coleborne, C., 2009, 'Families, insanity, and the psychiatric institution in Australia and New Zealand, 1860–1914', Health and History, 11(1), 65–82.

Coleborne, C., 2010, Madness in the family: insanity and institutions in the Australasian colonial world, 1860–1914, Palgrave MacMillan, Houndmills.

Gregory, I.N. & Ell, P.S., 2007, Historical GIS: Technologies, methodologies and scholarship, Cambridge University Press, Cambridge.

Jackson, W., 2013. Madness and marginality: the lives of Kenya's white insane. Manchester University Press, Manchester.

Travis, C., 2010, 'Abstract machine – Geographical Information Systems (GIS) for literary and cultural studies: 'Mapping Kavanagh'', International Journal of Humanities and Arts Computing, 4(1–2), 17–37.

Travis, C., 2013, 'GIS and History: Epistemologies, Re ections, and Considerations', in A. von Lünen & C. Travis (eds.), History and GIS: Epistemologies, Considerations and Re ections, pp. 173–193, Springer, Dordrecht.

Travis, C., 2014, 'Transcending the cube: translating GIScience time and space perspectives in a humanities GIS', International Journal of Geographical Information Science, 28(5), 1149–1164.

Blended Research on Blended Learning: Self-Re exive notes on a WITS Hybrid symposium about digital T&L

Christopher Fotheringham, Anita Virga, Brian Zuccala

Keywords: blended learning; blending symposium; hybrid curriculum development

Abstract

This paper may be understood as being meta-academic in nature, insofar as it discusses not so much a speci c piece of research on digital Teaching and Learning as, rather, a way of conducting, collecting and sharing digital T&L-related research.

The paper illustrates the comparatively new format of the 'blended symposium', in the form in which it is being implemented at Wits School of Literature, Languages and Media, as part of the University newly designed Digital Strategy. The talk lays at the methodological intersection of di erent scholarly areas, such as studies on blended learning, studies on tertiary education and/in the Global South, and environmental studies, and it unfolds in two sections. The rst part of the paper covers the main theoretical and pragmatic aspects of the Wits initiative, revolving around a phase of digital collection and circulation of pre-recorded video-contributions and a following moment of face-to-face workshopping.

The second part of the talk delves into the a ordances and limitations of such a model in relation to the tertiary education landscape in the Global South in general, and in the South African context in particular.

Digital humanities and academic libraries: Challenges and Implications for librarians and librarianship in South Africa

Oghenere Gabriel Salubi,
Ndakasharwa Muchaonyerwa

Keywords: Academic libraries, librarianship, digital humanities, South Africa

Abstract

From small technical and vocational education training (TVET) colleges to large university libraries, librarians continue to take on the challenges of digital humanities in di erent dimensions and a range of ways. The miscellany of expertise, understanding, and experience required to be displayed by librarians in the digital humanities continues to extend, expand, and develop. The limitations of digital tools, establishing good project development practices for digital humanities scholarship, managing data through the research lifecycle, tutoring multiple digital literacies (information, data, visual), and renegotiating the traditional librarian/faculty relationship from service orientated to full collaborator and partner are just some of the concerns facing librarians and librarianship in South Africa in the digital humanities domain. This paper re ects on some critical perspectives of planning and project management as well as the current challenges that inhibit academic librarians engaged within the academic community in digital humanities and implications for the librarianship profession in South Africa. The vast majority of librarians learn digital humanities on-the-job. In addressing professional development needs and evaluating the landscape of digital humanities and librarianship, the roles, scope of training and future training needs of professional library and information services in the digital humanities requires consideration. The available professional development opportunities, as well as institutional infrastructure requirements, also demand essential re ections. A close look at most of the library and information science curriculum in South Africa reveals that developing, planning and project management skills are not part of the traditional librarian education yet such

skills are essential for developing a successful digital humanities projects and are often learned on-the-job. While these skills should be reviewed into the curriculum, leveraging librarians' project management skills that technical, and administrative librarians already possess will enhance the successful operational skills necessary for the creation and e ective running of successful digital humanities projects. An incremental approach in supporting digital humanities projects and research needs through outreach, collaboration, and building technical skills should be considered.

The Exclusion of the Illiterate on Digital World: The Irony of Transformation.

Mdhluli Justice

Keywords: illiterate,digital, languages, decoloniality, exclusion

Abstract

There are countless number of digitally illiterate South African citizens who are eloquent and have advanced knowledge and deeper understandings on how di erent South African indigenous languages historically functioned successfully. They are also informative on how that history can be restored and still function in contemporary South Africa due to its undying relevance in solving social, economic, technology and political issues. They are in position of knowing how to use the correct dialects of the native language. The motive behind reference to these African indigenous legendary people draws its potency, according to the functionality of linguistics, that they are endowed with a rich linguistic and cultural background which cannot be mistaken nor be forgotten as part of historical consciousness and memory. The drastic development of technology and digital spaces has, with no doubt negatively impacted these people with social media platforms taking wave of human spaces which, in my believe, contributes to the distortion of the authenticity of African indigenous languages and literatures. This paper, utilizing an interdisciplinary approach believes that there is an element of exclusion between digital spaces and illiterate South Africans. In progress, holding the same belief, it has been observed that the isolation of these citizens from digital spaces has a negative impact on decoloniality of the indigenous languages project. Perhaps it should be reminded that one of the primary aims of decolonization is to provide access to success for all borne de members of the country; therefore, any domain, for example digital resources, that acts as a boundary slows down the intended results of the decoloniality project. It should also be highlighted that it is not the intention of this paper to criticize technology and digital resources instead it adds another dimension that can be used as a mirror or a re ection in a process of digital development and new inventions. It is also fundamental that as we are striving to connect all the o cial languages of South Africa to the world of technology and digital it has also be taken into consideration that proper application of the language is required. The South African constitution stipulates that all the eleven o cial languages should receive equal

treatment but truth must be told that some of the indigenous languages are still marginalized. The key concepts here; digital spaces and illiterate, although have a broader meaning, for the purpose of this research a narrowed de nition will be provided through the scrutinization of the existing research in the eld and world of academia. Another aspect that this paper intends to add is to call for the complete recognition of African indigenous languages through technology and its nitty gritty. It should be reminded that the fact that these indigenous languages and their theoretical approaches are used in the arena of academic inquiry is the proof that more theories that are epistemologically multidimensional in nature can be developed and implemented. This is a long overdue task that the country is facing, not only in South Africa, but the globe as well.

Umuntu ngumuntu ngabantu to umuntu ngumuntu ngeentu zakhe: The case of consumerism and materialism in a technological world

Diko Mlamli

Keywords: Ubuntu, Umntu, Consumerism, Materialism, Social Media

Abstract

This paper, through qualitative research methodology intends to critically provide perspectives on the transitioning from Umntu Ngumntu Ngabantu (you are who you are because of other people) to Umntu Ngumntu Ngeento Zakhe (you are who you are because of what you have). Over the centuries, materialism and consumerism have taken an enormous part of our human existence such that people have manipulated ubuntu – umntu ngumntu ngabantu. This has become habitual such that modernly, people value you based on what you have consumed materialistically instead of valuing who you are – a human being that is not di erent from another human being, after all we all equal in God's eyes. The radical development of technology or digital resources has de nitely a ected our identity. I would like to believe that the problem is what social media presents to us; the public. For example, instagram has been widely used to represent wealth and extreme consumption of expensive material such as gadgets, jewelries and so on. It should also be highlighted that the aim of this paper is not to criticize technology or the existence of digital resources but to highlight the signi cant impact these modes have had in African identity. Perhaps there is a need to re-search and re-think how technology, such as social media, can be advanced without subverting our identity and ubuntu.

Furthermore, umntu ngumntu ngeento zakhe does not only re ect on social media platforms only. This trend or attribute appears in our societies where people no longer value each other, there is no respect and there is no dignity amongst ourselves. People who are recognized and respected in our societies are those who have consumed more of materialistic belongings and properties. The question that one asks himself is, when and how did we get here? The new generation is engulfed by the celebrity stars, whom I would like to believe that they do not really embody real human existence instead there is this idea of consuming and consuming wealth.

Celebrities do not embody real human existence instead they embody perfection and that one cannot make a mistake. Probably this is the reason why most people opt to make arti cial looks such as skin bleaching, body chemical surgeries and plastic surgeries because you will only be valued as umntu when you look “perfect”. Everyday each person is exposed to all the riches, wealth, materialism and consumerism in social media. Perhaps, there is a need for the new generation to be taught and constantly reminded of their history and memory needs to be restored. Ukuxatyiswa komntu yaye iyinto yabaNtsundu futhi kungekho nto isisikhalazo kuloo nto, kuthe ke ngoku ka kwamaKoloniyali zaze zabhunyulwa ezo ntsika.

In conclusion, the sad and di cult thing in this is that most people die inside trying to maintain this idea of consuming material belongings and properties because they want to be recognized as human beings. Most people have died trying to maintain this obsession, trying to impress social media residents and society. It sorts of gives an impression that one disconnects themselves from who they really are. In simple terms people value reputation instead of character.

A quantitative comparison of isiZulu and isiXhosa named entity morphological structure.

Andiswa Bukula Rooweither Mabuya

Keywords: Nguni languages, named entities, agglutinative, morphological structure

Abstract

A quantitative comparison of isiZulu and isiXhosa named entity morphological structure.

IsiZulu and isiXhosa are Nguni languages classi ed in the South-eastern geographical zone (Guthrie, 1969; 1970), and are two of the o cial languages of South Africa. The two languages are the most widely spoken South African languages, with approximately 9 and 8 million mother-tongue speakers, respectively. In terms of natural language processing, however, both of these languages are considered under-resourced languages, with comparatively little data and resources available (Favre, et al., 2005).

Nguni languages are characterised by a rich agglutinating morphological structure, based on two principles, namely the nominal classi cation system, and the concordial agreement system (Bosch and Pretorius, 2009). This paper investigates the cross-linguistic similarities and dissimilarities between related languages o er in terms of morphological structure, speci cally for so-called named entities (NEs). NEs are words or phrases identifying rigid designators, such as the name of a person or an organisation, location, brand, product, etc., which references a speci c entity with a unique identi er (Abdallah et al, 2012). For the purpose of this research, the study is limited to three NE categories, person (PER), location (LOC), and organisation (ORG) (Eiselen, 2016).

The study uses quantitative analysis to evaluate the similarities and dissimilarities between the two languages, focusing on the relative frequency of the di erent classes, as

well as the a xes associated with the three classes. For this purpose, we evaluate the NCHLT named entity annotated corpora (Podile & Eiselen, 2016; Manzini & Eiselen, 2016), consisting of approximately 15,000 tokens annotated for one of the three NER classes. As an example, several in ectional variants found in isiZulu, either never, or very infrequently occur in isiXhosa. This is illustrated through the occurrence of ve variants in the isiZulu LOC class, ase-, base-,lase- nase-, wase- and yase-, that never appear in the isiXhosa LOC annotated data. In addition to the analysis of a xes, the quantitative analysis also looks at the variants in terms of capitalization, word length, singular and plural pre x frequencies, and hyphenation. According to Ndimande-Hlongwa (2012), both isiZulu and isiXhosa orthographies, hyphenation is used:

" when a numeral is preceded by an in ected pre x to join concords to Arabic numerals,

" to separate two vowels, or

" is inserted between the initial vowel pre x and an acronym.

In the data set that is being used for the analysis section of this paper there are several inconsistencies related to hyphenation. These inconsistencies are a concern because these languages are standard languages and standard languages should have notions of what correct and incorrect is, according to the isiXhosa and isiZulu National Language Boards (NLB). The paper will contribute to the discussions around computational textual studies in the eld of isiXhosa and isiZulu.

Data reference list

Podile, K., & Eiselen, R. 2016, NCHLT isiXhosa Named Entity Annotated Corpus, electronic dataset, viewed 28 November 2018, https://hdl.handle.net/20.500.12185/319

Manzini, A.N. & Eiselen, R. 2016, NCHLT isiZulu Named Entity Annotated Corpus, electronic dataset, viewed 28 November 2018, https://hdl.handle.net/20.500.12185/319

Reference list

Abdallah, S., Shaalan, K., and Muhammad, S. 2012. Integrating rule-based system with classi cation for Arabic named entity rec

Gami cation as an intervention tool in the teaching and learning of Setswana at an ODL Institution

Violet Pule

Keywords: Gami cation, Grammar, Setswana, Design, ODeL

Abstract

One of the signi cant aspects of learning grammar is knowing how to put words together in a meaningful order according to syntactic-semantic rules. Also, it plays an important part in improving listening, speaking, reading and writing skills. When one talks about grammar the major branches of linguistics emerge, namely phonology and phonetics, syntax, semantics, morphology, pragmatics as well as sociolinguistics. One of the main challenges of learning at distance

institutions is the absence of face to face interaction between a teacher or a lecturer and a student within the learning environment. Studies which have been conducted by Brindley & Paul (2004), Garrison & Shale (1990) and Lave & Wonger (1991) clearly points it out that students' development is determined by social interaction through problem-solving under the guidance of a teacher or in collaboration with capable peers. Devices such as cell phones, tablets and laptops become empirical in bridging the gap between students and their teachers or lecturers and vice versa. In addition, it also acts as a cognitive delivery tool in their learning journey. The interrogation should be how one can successfully learn the mentioned branches of linguistics at the absence of a lecturer and only rely on technology or digital resources. This research intends to design and evaluate a digital language learning application towards enhancing interaction and support in a distance learning context, speci cally learning Setswana grammar at a distance institution using mobile applications. Hopefully it will improve the idea of intellectualizing African indigenous languages in the so-called controlling domains (Maseko, 2011). It must be highlighted that the meaning of the term "linguistics" in this paper has been narrowed down to refer to the grammatical layout or aspect of Setswana language. The main research question for this study is how the use of gami cation can support the interaction of teaching and learning of Setswana grammar for distance educational environment for students to be able to acquire the relevant and required skills. The research will be framed within the two theoretical frameworks which are transactional theory and pedagogy theory. The research will use the design base research (DBR) which pragmatically employs qualitative and/or quantitative research methods (MacDonald, 2002) that are corresponding with the research questions and interrelated discussions throughout the research. The utilization of these two research methods; qualitative and quantitative, is motivated by the idea that one method has its own weaknesses, and in that situation Mkonto (2005) suggests that the other method becomes instrumental in uplifting the weaknesses of the other method. The signi cance of this study is the contribution it will make towards learning African languages through modern technologies in the arena of academic inquiry.

Word Frequency Distributions of African Languages

Trudie Strauss, Michael J. von Maltitz, Damián E. Blasi

Keywords: Word Frequency, Zipf's Law, African Languages, Language Diversity

Abstract

The languages of the world do not only di er from one another, but they also host much internal variation. This variation is expressed across many time scales: the split second reaction in the brain as languages are processed, the years over which children acquire their rst language, and the centuries over which languages change and evolve. The question then arises of how to compare, quantify and evaluate these diverse dimensions of variation. Yet most

linguistic units - such as communicative conventions, morphemes and syntactic rules - do not easily allow for large-scale comparisons. Fortunately, a seemingly simple property of languages, the distribution of the frequencies of its words, is not only largely comparable across language but it also plays a substantial role in the domains of language acquisition, processing and evolution. It has been known for a long time that word frequency distributions display similar statistical properties across languages, following a distribution known as Zipf's Law, which roughly corresponds to a power-law distribution. While this comparative and quantitative approach imposes some simplifying assumptions to the linguistic material, it provides some guarantees of unbiasedness against excessive theorising based on a handful of well studied, usually European, languages.

Twenty languages from three language families spoken across the African continent are considered in this study. Since it is not surprising that Zipf's Law also approximately holds for these languages, the word frequency distributions for these languages are investigated even further, and the deviation from Zipf's Law is explored. Several linguistically relevant measures are calculated from the frequency distributions and as such, each language is described in a multidimensional space. Findings show that some of the calculated measures exhibit structure in the deviation from Zipf's law and the linguistic interpretation of this structure is investigated.

Twenty languages from three language families spoken across the African continent are considered in this study. Since it is not surprising that Zipf's Law also approximately holds for these languages, the word frequency distributions for these languages are investigated even further, and the deviation from Zipf's Law is explored. Several linguistically relevant measures are calculated from the frequency distributions and as such, each language is described in a multidimensional space. Findings show that some of the calculated measures exhibit structure in the deviation from Zipf's law and the linguistic interpretation of this structure is investigated.

Technology and Identity: Ubugqirha in digital spaces

Pumela Chitani

Keywords: ubugqirha, indigenous knowledge, tradition, media, identity

Abstract

This paper seeks to aesthetically interrogate, through observational, descriptive, investigative and analytical approach, the impression that social media, such as television, creates by exposing cultural sensitivities such as ubugqirha (traditional healing) which is the primary traditional custom that will be scrutinized in this paper in relation to these domains. The motive behind reference to this indigenous traditional custom draws its strength from the conversations and criticisms that emerged from a South African lm, Inxeba that aired in 2017. The study and revisit into African indigenous knowledge systems as well as historical consciousness and memory is an overdue task facing scholarly critics. In addition, this paper seeks to reverse the misinterpretations that have been associated with ubugqirha which could be the possible reasons that contributed to, in my belief, the disrespect of the tradition. As the world, and South Africa in particular embarks on the transformation of its history and decolonization of spaces, it is empirical for this paper to rewrite what has been historically written by the "other" about this tradition.

Improving a Sesotho online dictionary.

Mmasibidi Setaka

Keywords: Lexicography, corpora, Sesotho, NCHLT Sesotho Text Corpora

Abstract

Technology has transformed the landscape of language studies, speci cally lexicography, which has bene ted immensely from its contribution to corpora (Nkomo, 2008). This o ers new opportunities to improve lexicography for African languages, and for Sesotho speci cally. Dictionaries produced prior to the availability of electronic corpora and large-scale corpus-based studies have often been criticized, hence Gouws (1990: 55) argues that "Unfortunately, the majority of these dictionaries are the products of limited e orts not re ecting a high standard of lexicographic achievement". However, e orts have been made to improve the quality of Northern Sotho dictionaries, for example Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete , but the use of technology for Sesotho has been limited. For this reason, it is important for Sesotho lexicographers to inherit the technological advantages now available to make their dictionaries more applicable for use in economic, educational, scienti c, and various other elds. In this study, various technologies are discussed that would improve the lexicographic process for Sesotho, by applying these techniques to an existing on-line dictionary for Sesotho, namely the Bukantswe online dictionary .

Electronic corpora are of paramount importance and Prinsloo and De Schryver

(2001) emphasize this when they state "... that more and more serious contemporary linguistic applications are based on electronic corpora [and] if African linguistics is to take its rightful place in the new millennium, the active compilation, querying and application of corpora should therefore become an absolute priority." The corpus era led to the development of improved digital lexicographic practices and allowed for speed, better article structures, improved examples of usage and similar advances.

One of the rst dictionaries to use corpora as a fundamental building block was the COBUILD dictionary , published in 1987. It revolutionized dictionaries for learners, completely changing approaches to dictionary-writing, and leading to a new generation of corpus-driven dictionaries and reference materials for English language learners. COBUILD served as an example for other dictionaries and it is in this vein that this study focuses on improving the quality of the current Bukantswe online dictionary (a bilingual Sesotho- English online dictionary) in terms of the treatment of words and richness of information contained in each entry. The purpose of this study is therefore to establish a way of improving Bukantswe using digital methods.

Electronic corpora transformed the way dictionaries have been compiled. The aim is thus to use similar methods that were used by dictionaries such as Macmillan , Collins , Oxford etc., to improve the Sesotho dictionary. Electronic corpora are also central to this study because "The rst task in compiling a corpus-based dictionary for any language is to build or to have access to a corpus" (De Schryver 2008:268). Corpora give access to huge amounts of data in a short period, which saves time, is quality based, and give lexicographers more information on what to include and exclude and on how to treat each article. It is therefore important that current Sesotho lexicographers take advantage of this ground-breaking innovation.

In this study we focus on the Bukantswe Online dictionary and check 1) inclusion and omission of words and 2) check the consistency of the treatment of words. This dictionary was last updated in 2008 and has limited entries (10 075), compared to the Collins dictionary, which has 725, 000 words, meanings and phrases. The NCHLT Sesotho Text Corpora (Puttkammer et al., 2014) is used to do frequency tests and also to extract concordance lines to assist in getting usage examples. This study shows that by using electronic corpora, it is relatively easy to extract frequently used words and lemmatise them. For example, it is through frequency studies that highly used words such as (frequencies given in brackets) latswa (54) 'taste', laya (67) 'advice', lebenkele (45) 'shop', lebese (131) 'milk', leeto (117) 'journey', letsopa 'clay', are included and treated in a dictionary because they are most likely to be looked for by users.

The following example shows concordance lines for kgasa 'crawl'. In this instance, examples of usage of the word 'kgasa' can be extracted and then put in the dictionary.

In addition to the use of corpora, the Fieldworks Language Explorer (FLEx) is explored because it is dictionary software that enables linguists to be highly productive when building a lexicon and

interlinearising texts. The software also has powerful bulk editing tools that can save a lexicographer hours of work, and allows for control of which elds and entries show up in a dictionary publication. FLEx also allows users to collaborate with colleagues located anywhere in the world.

It is therefore possible to improve the Bukantswe online dictionary through the use of digital methods. This paper will provide practical steps that can be taken when building online dictionaries for African languages through the use of digital methods and will also give input on what should be avoided to ensure quality dictionaries.

Data reference

Puttkammer, M.J., Schlemmer, M., Pienaar, W., & Bekker, R. 2014. NCHLT Sesotho Text Corpora. Electronic data set, North-West University; Centre for Text Technology (CTexT). http://hdl.handle.net/20.500.12185/336

References

Bukantswe online (2008). http://bukantswe.sesotho.org Date of access 30 Nov. 2018

Collins dictionary (2018). https://www.collinsdictionary.com/ Date of access 30 Nov. 2018

FieldWorks (2018). In FieldWorks Language Technology. Retrieved from https://software.sil.org/ eldworks/

Gouws, R.H. (1990). Information categories in dictionaries, with special reference to Southern Africa. In R.R.K. Hartmann (Ed.), Lexicography in Africa. Progress reports from the Dictionary Research Danie Prinsloo, Analysing words as a social enterprise, AustraLex 2015, page 15 Centre Workshop at Exeter, 24–26 March 1989: Exeter Linguistic Studies 15 (pp. 52–65). Exeter, England: University of Exeter Press

Macmillan dictionary (2018). https://www.macmillandictionary.com/ Date of access 30 Nov. 2018

Nkomo, D (2008) Towards a Theoretical Model for LSP Lexicography in Ndebele witrh Reference to a Dictionary of Lingistics and Literary Terms. Master of Philosophy. Stellenbosch University. South Africa

Oxford dictionaries (2018). https://www.oxforddictionaries.com/ Date of access 30 Nov. 2018

Prinsloo, D.J. and De Schryver, G.M (2001) The compilation of electronic corpora, with special reference to the African languages SOUTHERN AFRICAN LINGUISTICS AND APPLIED LANGUAGE STUDIES ISSN 1607–3614 Southern African Linguistics and Applied Language Studies 19: 111–13

Pukuntšutlhaloši ya Sesotho sa Leboa ka Inthanete (n.d.). https://africanlanguages.com/psl/ Date of Access: 30 Nov. 2018

Designing a Sesotho sa Leboa dialectal speech corpus

Dimakatso Mathe

Keywords: digitization, digital technology, mapping, dialects, speech corpus

Abstract

Designing a Sesotho sa Leboa dialectal speech corpus

Over the last decade, there have been signi cant e orts to extend available speech corpora for the South African languages, (Barnard et al., 2009; Badenhorst et al., 2011; de Vries et al., 2014). Although these resources are useful for the development of Human Language Technology applications, they are limited in their scope, speci cally to how di erent dialects in the South African languages are represented. This is despite the fact that many of the African languages are characterised by a high number of dialects with no orthography, nor published material. The focus of this study is the collection of a multi-dialect speech corpus for Sesotho sa Leboa, one of South Africa's o cial languages.

Gouws and Prinsloo (2004:21) acknowledge the predicament of lack of spoken data and the signi cance of such corpora for African languages with no written sources, however, logistics involved in such corpus collection are cited as major obstacles. This has negative implication on dialects of Sesotho sa Leboa as corpora continue to be built using standard language (Mojela, 2014), while very little work is done for the various dialects of the languages. Without a multi-dialect speech corpus, inquiry to explore the nature of dialects has limitations. Furthermore, development of Natural Language Processing tools, such as automatic speech recognition and text to speech systems in Sesotho sa Leboa are based on standard language (Badenhorst et al., 2011; Barnard, 2009) and are not suitable for processing dialectal variants of Sesotho sa Leboa. With this background in mind, the study seeks to propose a process for collecting, digitising, and visually representing a multi-dialect Sesotho sa Leboa speech corpus, and to demonstrate the signi cance of such an endeavour. It is vital that speech corpora of the dialects be collected and stored in such a way that it will be possible to do comparisons on the di erent dialects.

In addition to the corpus design, and data collection procedure, the study provides an overview of Sesotho sa Leboa dialects, with special attention to the geographic distribution of dialects, issues regarding their classi cation, and common challenges in relation to the dialects and the standardised form. Attention will focus on the methodology of collecting a multi-dialect speech corpus; selection of representative samples, means of digitising data, and the process of data representation through visual mapping of the corpus to make it accessible to the public in an easily understandable manner. The study will conclude by providing envisaged usage of the speech corpus and how such e orts contribute to enable research expansion in African languages in general, and under-resourced languages in particular.

References

Badenhorst, Jaco, Charl van Heerden, Marelie Davel, & Etienne Barnard. (2011).

"Collecting and evaluating speech recognition corpora for 11 South African languages." Language resources and evaluation, Vol. 45, No. 3. pp 289-309.

Barnard, E. Davel, M. & van Heerden, C. 2009. "ASR Corpus Design for Resource-Scarce Languages," in Proceedings of the 10th Annual Conference of the International Speech Communication Association (Interspeech), Brighton, United Kingdom, September, pp. 2847-2850.

de Vries, N.J. Davel, M.H. Badenhorst, J. Basson, D. de Wet, F. Barnard, E & de Waal, A. (2014). "A smartphone-based ASR data collection tool for under-resourced languages." Speech Communication, Vol 56, pp 119–131.

Gouws, R.H. & Prinsloo, D.J. (2005). Principles and Practice of South African Lexicography. Stellenbosch: SUN PReSS.

Mojela, V.M. (2013). "A Balanced and Representative Corpus: The E ects of Strict Corpus-based Dictionary." Lexikos, Vol. 23. pp 286-296.

Taljard, E. Faaß, G. Heid, U. & Prinsloo, D.J. (2008). "On the development of a tagset for Northern Sotho with special reference to the issue of standardisation." Literator, Vol. 29, No.1. pp 111-137.

Long Papers

Tracing the Development of Digital Humanities in Australia

Paul Arthur

Keywords: e-Research Infrastructure, Digital Resources, Cultural Data, International Engagement, Australasian Association for Digital Humanities

Abstract

This paper traces the development of digital humanities in Australia, with reference to major projects and events leading to the founding of the Australasian Association for Digital Humanities (aaDH). It discusses national exemplar projects, as well as signi cant activities and initiatives that formed a basis for the Australian eld. It outlines the history of the establishment of aaDH as a regional association, re ecting on its directions over the past decade, and describes the parallel development of large-scale infrastructure that has supported the eld's further growth. Looking back, it becomes clear that many projects and activities were pointing in the same direction without necessarily being linked directly, but as they progressed and connected with each other, the dispersed and disparate work across disciplines, institutions and sectors began to cohere with a common purpose—and from this con uence the Australasian Association for Digital Humanities (aaDH) was formed.

A range of large-scale digital projects and events provided the key context and impetus for the development of the digital humanities eld in Australia. Projects discussed in this paper include those of the Centre for Literary and Linguistic Computing at the University of Newcastle; the Archaeological Computing Laboratory (later Arts eResearch) at the University of Sydney; the eScholarship Research Centre at the University of Melbourne (and its predecessors); the Consortium for Research and Information Outreach (later Centre for Digital Humanities) at the Australian National University; 'AustLit'; 'AusStage'; 'Design and Art Australia Online'; 'Paradisec'; the 'Australian Dictionary of Biography'; and the 'Trove' digitisation and aggregation project of the National Library of Australia. As is the case in many parts of the world, such institutional activities tend to be associated with particular research centres or groups, and are typically collaborative enterprises.

The paper highlights the role of foundational projects and events for the development of the digital humanities eld in Australia, and in this context, it traces activities of the Australasian Association for Digital Humanities as a peak body representing the interests of researchers and connecting them with an international community. aaDH has been active for almost a decade and is now expanding its engagement with the next generation of scholar-practitioners in universities and the GLAM sector in Australasia. Meanwhile, since the association's establishment, the eld has expanded exponentially to become a major mode of connection between research communities operating in very di erent national contexts, serving as a means of global linkage, promoting and opening up possibilities for a higher degree of international collaboration.

Major projects have helped to determine priorities and interests for the eld as a whole and to foster communities of practice. In historical and disciplinary

terms, many can be linked directly back to the humanities computing tradition that predated and provided the foundation for digital humanities as it is now known. Leading projects have facilitated cooperation across geographical, cultural and disciplinary boundaries and integrated data in ways that enhance capacity, scale and accessibility. Through these characteristics they have been able to stimulate, generate and support new kinds of research activity. Smaller-scale projects can be equally innovative, and in the Australian context they have also contributed to the expansion of the digital humanities eld over time. However, the aim here is to highlight a selection of the most visible and active projects that can be regarded as iconic exemplars. They are large-scale collaborative initiatives that have been in uential in the eld, and through their high pro le and long reach, they have played a signi cant role in raising awareness of new digital approaches and possibilities. Linking the examples is a common theme: they all have the goal of capturing, preserving, building upon and articulating the richness of Australian culture and history for current and future generations.

References

Austlit (n. d.). Retrieved from https://www.austlit.edu.au

AusStage (2018). The Australian Live Performance Database. Retrieved from https://www.ausstage.edu.au

Australasian Association for Digital Humanities (n.d.). aaDH. Retrieved from https://aa-dh.org

Australian Dictionary of Biography (1966–). Australian Dictionary of Biography. Retrieved from http://adb.anu.edu.au

Design and Art Australia Online (n.d.). About the DAAO. Retrieved from https://www.daao.org.au

Paradisec (the Paci c and Regional Archive for Digital Sources in Endangered Cultures). (n.d.). Retrieved from http://www.paradisec.org.au

Trove (n.d.). Retrieved from https://trove.nla.gov.au

Multi-sensory experiences: site-speci c digital literature for and with persons with visual impairment

Franci Greyling

Keywords: site-speci c digital literature, multi-sensory garden, visual impairment, multi-modal texts, experiences

Abstract

Site-speci c digital literature – a form of digital literature combining multimodal texts with a speci c concrete site and requiring the reader's physical presence – o ers many new possibilities for creation, publication, reception, experience and interaction. With an increasing awareness of the e ects of digital technology on humans and the environment, site-speci c digital arts and literature can help to renew the relationship between human, place and time and create new relationships. The potential and value of site-speci c digital literature projects are con rmed by the

rst-hand experience with Byderhand – an interdisciplinary creative and research project aimed at investigating site-speci c digital literature through practice-based research and from various perspectives. Through the presentation of the Byderhand projects, the community is involved and empowered in various ways.

In 2017, the Byderhand team was invited to become involved in the development of a multi-sensory garden at a school for children with visual barriers. In this context, a multi-sensory environment is a dedicated space created for educational purposes and speci cally makes provision for students with special education needs, to provide a stimulating environment in a safe area (Hussein 2010). Site-speci c digital literature is particularly suitable to form part of such an enriched environment. Whereas the reader of a standard printed text is not space bound and call to mind the sensory modalities (sight, hearing, touch, smell and taste), readers of site-speci c digital literature nd themselves in a certain place and experience the environment concretely and sensorily. In addition to the aforementioned senses, the vestibular, proprioceptic and kinesthetic sensory systems are inherent to a person's body consciousness and spatial orientation. The reader furthermore participates in the reading experience through various performative actions. In this way, the literature, the speci c place, the sensory experience and the active participation of the reader combine into a concrete reading experience. Access to these potential multisensory and potentially engaging experiences must be mediated and accomplished in various ways. The author/creator of site-speci c digital literature (and especially the production team of such projects) must therefore consider various practical and pragmatic aspects, in addition to the content of the work. These aspects include the re ection on context and place, interface design, way nding elements and the performative action of the participant. A particular challenge of this project was to make the sight-speci c digital literature platform, Byderhand, accessible to blind and partially sighted persons.

Several institutions and individuals were involved in the development of the multi-sensory garden. The garden o ers learners a special multi-sensory experience: the plants in the garden have been selected with the focus on smell and touch; the paths are laid out with di erent textures; outdoor musical instruments made of recycled material invite the learners to play along and to make music, and a mural with idioms about sound, provides the nishing touches. The contribution of the Byderhand project to the multi-sensory garden includes 10 garden verses and children's verses (also by blind poets), supplemented with musical arrangements, translations and typographic animations. A combination of concrete and digital interfaces provides access to the work. The learners can now scan QR codes in the garden, either on their own or together with others, and read and listen to di erent versions of the garden verses on their cell phones.

Two related Byderhand projects have developed in conjunction with the multi-sensory garden. The Pioneer stories are a collection of stories, narratives and poems by learners, former learners and sta of the school. These digital texts have been placed

on the school grounds (at the places with which the narrative is concerned) and are also available in a local co ee shop. Karoo Gardens is a collection of environmental poems that will be available to be experienced at the planned Braille Trail in the Karoo Desert National Botanic Garden. Translations of the poems are available in English, German, Portuguese and isiXhosa, among others.

The multi-sensory garden o ers learners and sta at the school a stimulating environment where they can relax and explore. The garden is also used for various educational activities. Observation of the interaction in the garden and feedback on the project con rm that making site-speci c digital literature in such a multi-sensory environment accessible to persons with visual impairment, can contribute to an enriched and deepened experience of place and literature. The three projects o er a special opportunity to investigate the multi-sensory experiences and site-speci c digital literature in di erent contexts. Finally, the project illustrates that creative disciplines can practically contribute to enriched environments and the development of enabling technology.

Interface design and accessibility for site-speci c digital literature for people with low vision

Gustaf Tempelho

Keywords: Accessibility, Low Vision, Site-speci c Digital Literature, Interface Design, Collaborative Design

Abstract

This paper is a re ection on the challenges and problem solving involved in the design of an inclusive and accessible user interface for people with visual impairment (low vision and blind). The research and design processes included experimentation with both braille and digital interfaces. As a multimedia designer I am involved in an on-going site-speci c digital literature project, Byderhand (At hand). In 2017 a school for learners with visual barriers in the Western Cape approached us about the possibility of incorporating digital literature as part of an envisioned multi-sensory garden project at the school. This new context implied that we had to reconsider the design and functionality of the existing project's interface. Interface design for mobile screens is in many ways based on visual communication, resulting in the condensing and simplifying of information into understandable graphical representations. This becomes a concern when users do not have the capabilities to see what is displayed on a screen. It is therefore imperative for designers to gain an understanding of blind and visually impaired users' needs and requirements regarding their interaction with mobile technology.

During the development of this project we collaborated with various people who are visually impaired and/or provide a service to people with visual impairment. This collaborative and participatory approach was especially relevant during the user research and evaluation phases of the interface design proses. The collaboration lead to a better awareness and a deeper level of thinking regarding our project goals. As a multimedia designer I had to overcome two main obstacles to ensure the usability of the Byderhand platform. The rst was to enable users with visual impairment to gain access to the digital interface via QR codes. This required a solution, as the scanning of QR codes relies mainly on a user's visual capabilities. The second goal was to design an inclusive navigational interface to allow users with di erent levels of vision to experience the multimodal texts published on the platform.

The development process, which spanned over the course of twelve months, entailed extensive user research, prototyping, evaluation and improvement of the physical and digital interfaces. The multi-sensory garden was unveiled on 30 August 2018. Learners can now interact and experience the multi-modal site-speci c digital literature on their school grounds. Parts of the project are also available for the general public, in the local co ee shop at Innovation for the Blind. The new Byderhand project was well received and is also used as an educative tool to stimulate learners and expose them to new experiences of literature. A further expansion of the Byderhand project is currently being developed as part of the Braille Trail in the Karoo National Botanical Garden. With more exploration, the design solutions used in this project could o er new possibilities for accessible spaces as well as alternative publication solutions for people with visual impairment.

2018 Technology Audit of Language Resource availability in South Africa

Carmen Moors, Karen Calteaux, Ilana Wilken

Keywords: Language resources, human language technologies

Abstract

Summary

This paper presents the approach taken to conduct an audit of human language technology (HLT) resources in South Africa and gives an overview of the ndings of this technology audit. Technology audits are a worldwide phenomenon and aim to provide insight into the technology landscape in a particular country or eld. The 2018 HLT Audit was commissioned by the South African Centre for Digital Language Resources (SADiLaR) and provides a framework for evidence-based decision-making on investment in local HLT resource development.

Background

Researchers, educators, developers, service providers and funders need a roadmap to enable them to decide where to concentrate their e orts in order to give a maximum

push to the development of a particular field, and to know what is available to enable further technology development and research. Technology audits are an important instrument to provide such a roadmap. Technology audits can also play a significant role in surfacing information which can be used by researchers, policy-makers and funders alike to build a country's research and development system of innovation towards increasing its competitiveness, contributing to its economy and bridging the digital divide.

Technology audits aim to identify, assess and catalogue technologies according to different criteria, ranging from categories of technologies, maturity of technologies, competitive position, location in the supply chain, and levels of competencies, through to the impact of technologies [1]. By employing a mapping technique known as a technology matrix, technology audits can provide an overview of the related technology landscape in a company, market or country.

2018 HLT Audit

The HLT Audit undertaken in 2017/2018 aimed to update our knowledge of the available HLT resources in South Africa, and replicate the 2009 HLT Audit [2] in order to facilitate a comparison with previous data.

The design and development of the 2018 Audit tool involved extensive research into past and current related audits and methodologies. The experts who participated in this process assisted in creating a simplified and modernised design for collecting information on existing HLT and other language resources. The design was implemented in an online tool as method to collect the data. Both the design and the resultant tool can be re-used (with minimal effort) to design future audits (if required) and continually capture HLT resources as these become available.

In addition to the datasets collected during the 2009 and 2018 HLT Audits, we obtained a dataset from the Resource Management Agency (the government-appointed custodians of HLT resources in the country and the precursor to the newly established SADiLaR) [3] containing the resources which had been added subsequent to the 2009 Audit. We therefore present an analysis of three datasets of HLT resources in this paper. We term these the 2009 dataset (emanating from the 2009 HLT Audit), the 2014 dataset (resources added between the two HLT Audits), and the 2018 dataset (emanating from the 2018 HLT Audit).

We match the resource types across all datasets and represent the number of resources in each language in each resource type across all datasets. A comparison of the availability and accessibility of these language resources across languages will be presented.

The comparison of the three datasets indicates that while significant progress has been made since 2009 to develop additional resources across more languages, and to develop cutting-edge resources (as we see from the types of resources added in 2018) and language independent resources, the more marginalised indigenous languages (particularly Xitsonga, Tshivenda, and isiNdebele), remain severely under-resourced. While there are more text resource types than speech resource types

across all the languages, there are still many resources that need to be developed in both text and speech.

Future work includes addressing the current challenges with the online tool, particularly the functionality to capture several similar resources with minimal e ort. Further work includes implementing a system(s) to ensure that HLT resources (and other language resources) are continually submitted to SADiLaR as these become available. Raising awareness on the bene ts of contributing to the body of knowledge and making resources available to others for further research and development will require focused attention.

Conclusion

The establishment of the South African Centre for Digital Language Resources (SADiLaR) marks a new era in language resource management in South Africa. SADiLaR presents a formal, government-funded, longer term (at least 10 years) infrastructure for curating and managing language resources. A formal process for submitting HLT resources and continually updating the database of HLT resources, is now possible. The current thinking on how to e ect this, is to follow a similar procedure to that implemented for the LRE Map in Europe, namely requiring researchers to submit resources mentioned in research papers and articles to SADiLaR before publication of such articles and papers. To this end, the online survey developed in this project (as reported on in a related paper) will be hosted by SADiLaR in future. In this manner, we aim to create a sustainable system for capturing, curating and distributing HLT resources in South Africa.

References

[1] U. Bross, "Technology audit as a policy instrument to improve innovations and industrial competitiveness in countries in transition," Innovation: The European Journal of Social Science Research, vol. 12, no. 3, pp. 397–412, 1999.

[2] A. Sharma Grover, "Technology Audit: The State of Human Language Technologies R&D in South Africa," Master's thesis, University of Pretoria, South Africa, 2009.

[3] South African Centre for Digital Language Resources (SADiLaR), 2018. [Online]. Available: https://www.sadilar.org/

Digital humanities meets sensory ethnography: analysing multisensory experiences through digital resources.

Jenni Lauwrens

Keywords: digital research methods, multisensoriality, ethnography, place

Abstract

This presentation describes and re ects critically on a learning experience that centered on a group of graduate students' multisensory, embodied and emplaced interactions in a public space. The project speci cally sought to explore new ways of doing visual research in the humanities in a digital era. The aim was to investigate how digital research methods (a tricky term that is explored in more detail in the presentation) can be utilised to create and analyse the data produced through questions asked from within the eld of the humanities and then disseminate the research ndings in a way that extends the 'creative possibilities' of digital technologies, as described by Schreibman, Siemens & Unsworth. (2016:17), for research in visual culture studies. At the same time, the project aimed to introduce graduate students enrolled in the Honours programme in Visual Studies at the University of Pretoria (UP) to digital research methods.

The project used digital research methods and tools to document, describe and interpret the students' experiences and then publish the results in a digital format. Mobile technologies were utilized to gather data and digital tools were used to construct a digital archive comprising audio-visual material collected during the data gathering phase. Thereafter, this data was analysed and the ndings were presented in an interactive, multimedia, hypermedia online platform. Scalar, which is an interactive digital platform that allows users to author and publish digital scholarship online, was used for this purpose. By taking an experimental approach, traditional research approaches took a digital turn in an e ort to grapple with the intersection of digital humanities and sensory ethnography.

The ndings showed that conducting research using digital resources can enhance collaboration between research participants. Furthermore, the collaborative digital working environment heightened empathic understandings of each participant's experiences giving rise to deeper meanings being generated about the ways in which multisensory experience produces place. Finally, the representation of the participants' experiences using digital resources rendered this new knowledge accessible to a wide audience.

The Importance of Digital Literacy and Online Research in HEIs

Mluleki Siguntu

Keywords: Digital Literacy, Online Research, HEIs

Abstract

The world is changing. We live in the ever changing world of Technology and everything is changing very fast. Research in the information age is no longer the issue of searching traditional sources as information is available in various di erent

formats i.e. ebooks, e-journals, research databases and online searches. This raises the need to facilitate Digital Literacy and prioritize Online Research. Digital literacy is the ability to access networked computer resources and use them e ectively (Gilster, 1997).

The Digital Age Society calls for all people to Become Digital Literate.

Online research is a research method that involves the collection of information from the internet. With the advent of the internet, the traditional pen-and-paper research techniques have taken a backseat and made room for online research. These raises the need to recognize the importance of online research. Online research is perhaps the most obvious but also the most di cult of research methods. (Tristram Hooley, 2011).

The skills of the digitally literacy and Online Research are becoming as necessary as a driver's license.

DL and online Research skill are considered to be a very important skill to possess in the rst world. Addressing the 21st century skill of digital society, digital literacy and Online Research skills are very important for students and sta .

Some attributes of include understanding how to use web browsers, Research Databases, Search engines, E-books, E-Journals, email, social media, wiki, blogs, Photoshop, Power point, Microsoft word and Excel, etc. to showcase learning. Evaluating online resources for accuracy/trust-worthiness of information and for e ective modern teaching and learning (cited, 2011).

This is of critical importance in the university setting where teaching approaches emphasize independence on the part of the student in accessing and synthesizing information from di erent sources in the creation of new knowledge (Kavulya, 2003).

The abilities to e ectively use and create technology to solve complex problems are the new and essential literacy skills of the twenty- rst century. Digital Literacy and Online Research are important and needed to all disciplines/ elds, all learning environments and to all levels of education.

The main objective of this paper is to raise awareness of the importance of digital literacy and Online Research. It looks brie y at where we come from and how far we have developed. It looks at reliable and unreliable online information sources.

This paper is largely informed by Digital Literacy and Computer Science standards in Higher education which provides framework for assessing an information literate individual. It is also informed by literature review of scholarly articles, books and Internet Information sources.

Enhancing the African Wordnets with existing lexicons

Sonja Bosch, Marissa Griesel

Keywords: African wordnet; multilingual lexicons; fast-tracking

Abstract

The African Wordnet Project (AWN) (see Griesel & Bosch, 2014 or https://africanwordnet.wordpress.com for a detailed introduction) has grown considerably since

the first workshop to introduce the concept to the South African linguistic community in 2007. The AWN currently includes synsets in seven languages, namely Setswana, isiZulu, isiXhosa, Tshivenda, Sesotho sa Leboa, Siswati and Sesotho. A proposal submitted to the South African Centre for Digital Language Resources (SADiLaR) to add Xitsonga and isiNdebele has also been approved and work on this will commence early in 2019. The manual development of the African wordnets follows the expand model (Vossen, 1998) and is based on the English Princeton WordNet (PWN) (Fellbaum, 1998).

Developing rich and hierarchically structured wordnets such as these for under resourced languages is labour intensive and costly. It is therefore important that the synsets represented in the AWN be an accurate representation of the most frequently used concepts, as this will ensure their usefulness for a variety of tasks within the South African Digital Humanities. If the wordnets were, for instance, used in the development of a computer assisted language learning environment, one would expect synsets for various basic concepts like body parts, items in and around the house, etc. A more technically advanced application such as information retrieval or machine translation would require the wordnets to cover different synonyms and hyponyms to ensure a wide coverage of actual language usage.

With this in mind, the AWN development team looked towards the SIL Comparative African Wordlist (SILCAWL; see https://www.eva.mpg.de/lingua/tools-at-lingboard/pdf/Snider_silewp2006-005.pdf). This comparative wordlist contains 1 700 glosses in English and French, categorised in 12 domains such as 'man's physical being', 'human civilisation' and 'plants'. The AWN team saw the value of such a list in further development of the African wordnets and as a means to standardise the content thereof over all languages. As a first step, the AWN team added links to the corresponding PWN entries to the SILCAWL glosses and extracted definitions as well as usage examples for the entries from the PWN. Since not all entries in PWN had this additional information, an English lexicographer was tasked with creating new definitions and/or usage examples to fill in these gaps and to make sure that the content would be lexicalised within the South African environment. The result was an initial database of 1 000 English terms with a definition and usage example each, linked to corresponding synsets in PWN. To fast-track development of the AWN and reduce the need for specialised linguistic knowledge in development of wordnet content, this database was next translated into the seven languages included in the AWN thus far. Translators lexicalised the content (so-called 'transcreation') to model the culturally appropriate real world usage for their target language. As a last step, the data was added to the complete AWN before linguistic experts confirmed the semantic links of the new synsets within the larger scope of the wordnets. This resulted in an additional 7 000 new synsets, complete with usage example and definition, added to the AWN, but more notably in a subset of synsets that are present in all seven languages. This not only enhances the usability of the AWN in direct comparative studies, but also in

natural language processing between the di erent South African languages.

In this presentation, we will show some interesting examples from this multilingual dataset as well as some applications of the AWN in the language learning and translation elds. Some suggestions for future application of the wordnets in the South African Digital Humanities environment will also be discussed.

References

Fellbaum, C., (ed), 1998. Wordnet: An electronic lexical database. The MIT Press, Cambridge, Mass. ISBN 978-0262-06-197-1

Griesel, M. & Bosch, S. 2014. Taking stock of the African Wordnet Project: 5 years of development. In Orav, H., Fellbaum, C. & Vossen, P. (eds): Proceedings of the 7th Global WordNet Conference 2014 (GWC2014), pp. 148-153, Tartu, Estonia, ISBN 978-9949-32-492-7.

Vossen, P. 1998. EuroWordNet: A multilingual database with lexical semantic networks. Kluwer Academic, Dordrecht. ISBN 0-7923-5295-5

Corpus cleaning strategies for African Language texts

D.J. Prinsloo, Elsabé Taljard

Keywords: corpus compilation, dirty corpora, African language texts, strategies for corpus cleaning

Abstract

Taken at face value the issue of 'clean' versus 'dirty/raw' and 'noisy' corpora seems to be quite simplistic – if the corpus is dirty, clean it, and save the clean copy for use by corpus query programs and corpus manipulation tools. In reality, however, cleaning a corpus is much more problematic than meets the eye. What is a dirty corpus and what constitutes a clean corpus? In fact, clean corpora is a relative concept in terms of e.g. granularity, that is, how clean should the corpus be. Atkins and Rundell (2008:93) state that lexicographers "prefer size to granularity" and "high volumes of data with the occasional bit of noise" instead of "very 'clean', carefully annotated data in much smaller quantities". Noise in text is de ned by Knoblock et al. (2007) as 'any kind of di erence between the surface form of a coded representation of the text and the intended, correct or original text'. Which version or versions of a particular text should be saved for querying and preservation purposes? Is a corpus clean when it matches the original document or does a clean corpus also presume (a) corrections of mistakes made by the author, (b) indication of circum exes, diacritics and tonal patterns, (c) the removal of foreign words and paragraphs and (d) omission of data irrelevant to the speci c intended application, (cf. Guthrie et al.), and speci cally relevant for our purposes, (e) correction of scanning errors? These questions can hardly be answered unless the purpose of the corpus is clearly de ned. Automatic sourcing of for example a special purpose corpus from the web can introduce sampling errors, i.e. retrieve documents that are not relevant to the particular research question and which could in the end introduce bias into the research results, cf. Kantner and Kutter, 2011.

Most of the current research on corpus cleaning report on cleaning of web- sourced corpora, cf. Hofmann & Weerkamp (2007). Web-based corpora usually consists of web pages, i.e. documents that are usually marked-up using HTML. As Hofmann and Weerkamp (2007) point out, such les contain navigational structures, e.g. menu's, headers such as logos and breadcrumbs, footers consisting of copyright notices and dates, and sometimes also advertisements. These unwanted pieces of text generate noise which makes processing di cult. The corpora that we report on originate from a variety of sources – some are web-based, some have been sourced in word processed format, but the bulk of the material have their origin in hard copy documents that were processed by means of OCR scanning.

This paper re ects on corpus cleaning activities for African language texts in the South African Centre for Digital Language Resources (SADiLaR) performed by the University of Pretoria's node for digitisation. Digitised data are used to build general purpose corpora, which are intended for utilization in multiple projects such as machine translation, writing assistants, spelling and grammar checkers, lexico-graphy, etc. for the African languages. Speci c attention will be given to types of errors and noise in these corpora such as typing and scanning errors and di erent types of repetitions, e.g. recurring material such as automatically generated text blocks, navigation bars, page headers, etc., cf. boilerplate repetitions, (Baroni and Kilgarri (2006), Evert (2008): NCLEANER). Manual, automated and semi-automated strategies to clean the corpora will be suggested and illustrated. It needs to be kept in mind that these languages are lesser-resourced in terms of both human expertise and computational tools needed for corpus cleaning. Issues such as incorrect spelling, old spelling, word division problems, grammatical errors, incorrect capitalization, etc.

References

Atkins, B.T. Sue and Rundell, Michael. 2008. The Oxford Guide to Practical Lexicography. Oxford, New York: Oxford University Press.

Baroni, Marco and Kilgarri , Adam. Large linguistically processed Web corpora for multiple languages. http://www.aclweb.org/anthology-new/E/E06/E06-2001.pdf

Evert, Stefan. A lightweight and e cient tool for cleaning Web pages. http://www.lrec-conf.org/proceedings/lrec2008/pdf/885_paper.pdf

Guthrie, David, Guthrie, Louise and Wilks, Yorick. An Unsupervised Approach for the Detection of Outliers in Corpora. http://www.lrec-conf.org/proceedings/lrec2008/pdf/866_paper.pdf

Hofmann, K and Weerkamp, W. 2007. Web corpus cleaning using Content and Structure. In Building and Exploring Web Corpora. Proceedings of the 3rd Web as Corpus Workshop.

Kantner, C and Kutter, A. 2011. How to get rid of the noise in the corpus: cleaning large samples of digital newspaper texts. International Relations Online Working Papers Series, 2011/12.

Knoblock Craig, Lopresti, Daniel and Shourya Roy and Venkata Subramaniam L.

2007. Introduction to special issue on noisy text analytics. IJDAR (2007) 10:127-128

NCLEANER http://webascorpus.sf.net

Commerce v. Scholarship: Digital Music Editions in South Africa

Christine Lucia, Zayne Upton

Keywords: Digital, Critical, Music, Commerce, Scholarship

Abstract

The website, www.african-composers-edition.co.za was developed in 2013 in order to present a new critical edition of works by southern African composer Joshua Pulumo Mohapeloa (1908-82), providing scholarly information on and samples of his music as well as selling 145 downloadable vocal scores and 31 audio tracks as a commercial venture set up in collaboration with the heirs of his estate. This 'static' website is currently being upgraded in order to create the nested layers of information, archival material or the interactive possibilities available on many digital critical editions globally, and provide a more useful service to performers, educators, and scholars of South African music. Two new editions in progress on the new site are of music by Surendran Reddy and Michael Moerane. Developing digital music editions raises many issues, among them the di erence between publishing, editing, scholarship, archiving, and commerce. Digital online editions not only provide greater social access to music and its histories but also generate income. Commerce is however conventionally seen as belonging in a realm outside scholarship. Does creating a digital online critical edition 'count' as research, especially if it sells?

The Importance of Fair and Transparent Machine Learning Practices for Digital Humanities

Karabo Maiyane, Emma Ruttkamp-Bloem

Keywords: machine learning, arti cial intelligence, responsibility, morality

Abstract

In the eld of 'critical machine learning', issues concerning fairness, accountability and transparency of machine learning practices are considered. The concern in this context is to determine and raise awareness of the e ects of knowledge symbolisation, classi cation and generalisation on the historical, political and social conditions of our lives.

In our talk we want to highlight the importance of believing responsibly. We will unpack William Cli ord's three motivations in support of the argument that believing responsibly is a moral obligation - 1) our beliefs in uence our actions; 2) poor practices of belief formation make for careless irresponsible believers; 3) we have a moral responsibility not to contaminate the beliefs of others. On the basis of Cli ord's work we will argue that 1) keeping in mind that it is wrong always to believe anything based on insu cient evidence is paramount in our era of digital access to research and big data, and 2) that decision making systems – i.e. machine learning models – need to be

used and their predictions interpreted responsibly.

Education 4.0: Making the Internet of Things (IoT) Relevant with Arduino in Preparation for the Fourth Industrial Revolution

Sean Kruger

Keywords: Education, Innovation, Internet of Things, Makerspace, Industry 4.0

Abstract

The Fourth Industrial Revolution, also referred to as Industry 4.0, surmises that we are amidst a wave of change that will ultimately alter the foundations of how we live. This change will occur through the adoption of technologies that in uences human life and behaviour. This change has been brought about due to the array of technologies which form part of Industry 4.0, that have varying applications and usages, including arti cial intelligence, augmented reality, autonomous vehicles, biotechnology, 3D printing, big data, data visualisation, internet of things, nanotechnology and quantum computing. The major di erence between this and other industry revolutions, is the pace at which it is occurring, and the impact it will have on technological ecosystems and system integrations. By altering the way people work, live and relate to one another, human interactions are changing. In the eld of digital humanities then, vast areas are being impacted with this movement, and if well leveraged, can unlock countless areas of research and practical application. What then are ways in which education can be updated to respond to the changing needs required of Industry 4.0 and enable new possibilities?

One area of Industry 4.0 that can be used to update academics in a practical way is the Internet of Things (IoT), which links physical infrastructure to computational capabilities. Their applications are almost limitless, irrespective of discipline, especially in the acquisition of data in areas that were either not possible to reach or due to other resource restraints. The advent and improved access to sensors, sensor networks, near frequency contact (NFC), radio-frequency identi cation (RFID), real time location system (RTLS) and wireless technology have opened up all these new opportunities that allow devices to monitor and sense the physical environments inputs to allow interaction between the digital and physical realm (Lee et al. 2015, 2017; Jun and Jing 2018).

By o ering Arduino training and kits, the Library Makerspace at the University of Pretoria is supporting Education 4.0 and driving research with the support of IoT devices. The key in this, is making the applications relevant through practical examples as well as guidance to researchers who have not yet seen the application within their speci c scope or eld of expertise.

It was found that by o ering training in a neutral area such as the Makerspace, positioned in the Library, academics have begun to apply IoT devices irrespective of discipline, where so many limitations can now be addressed, broadening areas of research. In practise then, cross disciplinary research has been further enabled with technology, including that which forms part of the digital humanities.

Thus, the purpose of this presentation will be to showcase the application of technology using Arduino kits to enable Education 4.0. This will be done by demonstrating what the technology does, and what outcomes it has achieved, especially to those relating to the digital humanities elds such as psychology, communication, linguistics and visual arts.

Connecting di erent "dots" to enhance learning

Liezel Blomerus

Keywords: University Access Programme, Personal Learning Environment, Conceptual framework, academic learning support

Abstract

In 2017 a conceptual framework for the o ering of academic learning support that enhance and develop a Personal Learning Environment has been developed for Anthropology modules presented in an Access Programme, at the University of the Free State. This conceptual framework was developed in relation to literature that suggested the concept of the Personal Learning Environment (PLE) to be an emerging technology based on the induced results of practices and challenges linked to social media and that it o ers great potential as a pedagogical approach, as well as an attempt to answer context speci c challenges faced by the Extended Programme that were narrowed down to issues related to under preparedness (First generation students, hours outside of class not utilized or accounted for and disruptive mindsets); the epistemological gap between secondary and tertiary education; and the shortfalls of pedagogical interventions.

The framework suggested a renegotiated and re-imagined understanding of o ering academic learning support that: 1) Tap into the natural learning setting of a student; 2) That cater for and make allowances for the PLE in the planning and implementation of academic learning support; and 3) That aims to foster, enhance and develop a student's PLE. This intervention rested on the presumption that this would cultivate and encourage personal development and empowerment, life-long and self-regulated learning, knowledge management and student control that would enable a student to deliberately structure and further design and develop their own personal learning environment.

The framework contented that the PLE, and not ICT in itself, allow for pedagogical a ordances. It centre and shape (through its input) pedagogical intervention that still remain aligned with the formal educational process and context. It then becomes the output of the interventions in the academic learning support environment as a self-embodied tool of a student that better engagement, enhancement and extension as well as personal development and empowerment. Synergy, uidity and interdependency therefore becomes the describing characteristic of the interaction between the formal academic learning support environment and the personal learning environment.

In line with the above the framework proposed that learning are initiated, enhanced and extended by three components: the individual or student, learning community spaces and the ICT

component. Learning are initiated by a students' engagement in learning opportunities and experiences within communities of practice. The communities of practice cultivate, foster and guide learning speci cally through its emphasis on translearning and heavy switching. The ICT component allow for learning to be extended through its emphasis on learning "from" and "with" technology approach. The learning process come full circle with the achievement of curriculum requirements, the ability to connect di erent dots and to transcend boundaries and the realisation of personal development and empowerment.

Against this background, the purpose of this paper is to discuss thematic themes which had emerged through the process of self-re exive practitioner inquiry in relation to some aspects of the proposed conceptual framework which were informally tested in 2018 within the University Access Programme. The above are informed by an action based research method that utilized qualitative data collection techniques and a thematic analyses approach.

Possible results may indicate suggestions on alterations needed for the conceptual framework to be applicable in a context were resources are in most instances limited, where students are not as digitally savvy as millennial student literature portray and where attributes and exit/transitional outcomes are framed around a very speci c embodied student that need intensive sets of academic support for undergraduate degree entrance.

A New Approach to the Dating of English Literary Texts: An application to the works of H Rider Haggard

Adrian Ryan Liliana Carrick-Tappeiner

Keywords: chronology, seriation, stylometry

Abstracts

While authorship identi cation using stylometry is a research problem that has attracted considerable academic interest, there has been relatively little work published on the problem of stylometric dating of literary works. In this paper we take two small and tentative steps toward the development of a computational toolkit for the study of chronology in literature. In particular, we present two stylometric methods to facilitate the construction of literary chronologies, one relative and one chronometric. In either case, we demonstrate the results of using two di erent feature sets, one derived from word co-occurrences and one from word frequencies. To assess the e cacy, we tested the method on the oeuvre of the British-South African author, H Rider Haggard. Haggard's work is particularly appropriate for testing since all his works are dated, he wrote quickly and 'automatically' meaning that his style is not heavily in uenced by editorial concerns, and nally because a large number of his works are freely available on Project Gutenberg.

The rst of the two methods takes its inspiration from relative chronological techniques in archaeology. Relative dating methods do not attempt directly to predict a date for a given subject, but instead to place a group of subjects in chronological

order. In archaeology the subjects are usually contexts or assemblages of artefacts, but in our case they are works of literature. The speci c method we propose is analogous to contextual seriation, which assumes that chronological change is, on the whole, smooth and gradual, and that the sequence that is most likely to be chronologically correct is that in which change between proximal subjects is minimised. In archaeology, the subject is usually described by a set of binary features (representing the presence or non-presence of speci c attributes) or categorical features, and they are arranged in order using multiple correspondence analysis (MCA) or similar methods.

Since the stylometric feature spaces we use are continuous rather than binary or categorical, MCA is an inappropriate tool. Instead we cast the problem of minimising change between successive subjects as a minimum Hamiltonian path problem for a complete weighted graph, in which each of the vertices represents a text, and the weights of an edge connecting two vertices represents the Euclidean distance between the two corresponding texts in the feature space. Such problems are, in terms of complexity, similar to the Travelling Salesman problem, and are not yet (and may never be) solvable in polynomial time. To obviate this problem, we use an o -the-shelf approximation to provide a good but sub-optimal solution. The results show that while it predicts sequences which correlate with the true chronological sequence with a probability signi cantly better than chance, it is not yet a viable tool.

The second approach we present is more robust, and simply employs linear regression to predict the date of a subject. For this purpose, we employ Principal Components Analysis on the whitened data, before applying the regression to the fewest principal components that describe over 90% of the variance. Using this technique, we achieve a high correlation, estimated via leave-one-out cross validation.

Over the course of the paper we will motivate the two methods and describe them in detail, explain the construction of the feature spaces on which they are tested, and the approach we have taken to assess the e cacy of the methods, before we present the results themselves. Finally, we draw some concluding observations and make some suggestions about how to proceed with further research. In particular, we consider whether the relative dating approach might still be useful if combined with the chronometric, and we consider what metrics, beside correlation, might be used to assess the e cacy of similar methods moving forward.

Developing parallel corpora for machine learning

Langa Khumalo

Keywords: Policy, Intellectualization, Data, Reusability, Parallel Corpus, Machine Learning

Abstract

The University of KwaZulu-Natal (henceforth UKZN) has a language policy that recognizes English and isiZulu as the two o cial languages for administration, teaching and learning, research and innovation. This has led to a program to

cultivate and intellectualize isiZulu so that it can be used in all spheres (Khumalo 2016, 2017). This has resulted in inter alia the production of massive textual data in both English and isiZulu. Most data produced pertain to manually produced translation and interpretation between the two languages. The increased demand for language services has led to the development of innovative digital solutions for isiZulu (Keet and Khumalo 2017). It is argued here that corpora are an important resource. They are an important precursor to the development of data-driven computational tools such as the spellchecker, automated machine translators and the lexicons. In the academy learner corpora are an important enabler in second language learning particularly in error detection. It is notable that every human activity, academic or social, simplex or complex, cultural, religious or scienti c produces discourse or text. It will be argued that corpora are representative of this complex matrix, and therefore provide invaluable evidence beyond comparison. Specialized corpora have recently been used to monitor disease surveillance (Brownstein, Freifeld and Mado 2009) and customer sentiment analysis in business (Pak and Paroubek 2010).

A parallel corpus is a collection of texts that are translated from one language L1 (the source language) to the other language L2 or languages LX (the target language(s)). In the case of two languages such as obtains at UKZN, one corpus is an exact translation of the other corpus. A Parallel Corpus thus typically contains data from two languages. In 2017 UKZN initiated a process to create a bilingual parallel corpus of English and isiZulu. The imperative to create the English-IsiZulu Parallel Corpus (EIPC) was inspired by the stated massive textual production in the two languages, as a result of conformity to the language policy, which stipulates that whenever possible administrative and academic information must be made available in the University's two o cial languages. The EIPC currently has 300 les of parallel text in both English and isiZulu. It is on the basis of the EIPC that a Data Driven Machine Translation (DDMT) approach will be employed to build a machine translation tool. The DDMT approach uses the data theory to and is based on the corpus such as the EIPC. For an under-resourced language such as isiZulu, the exigency exist to create such a tool in order to automate the translations between the two languages since human translation cannot cater for translation demand (Kituku et.al. 2016:1). We provide a test case for the EIPC.

Section 1 brie y describes the UKZN language policy in the context of the constitutional imperative to develop African languages in order to "[…] achieve parity" with the English language. Section 2 outlines data generation and processing at UKZN. Section 3 discusses digital methodologies in data storage for reusability. We describe the EIPC in Section 4 and discuss the test case in Section 5. We conclude in Section 6.

Aging and creativity. A Stylometric Analysis of Late Style in Johann Wolfgang Goethe's, Robert Musil's, and Franz Kafka's Works

Massimo Salgaro, Simone Rebora

Keywords: Late style, Stylometry, Johann Wolfgang Goethe, Robert Musil, Franz Kafka

Abstract

As many recent publications attest (Leeder 2015, Zanetti 2012, Seidler 2010, Said 2006), "Late Style" has become a fashionable concept in literary theory. The term expresses the idea that the art produced during the nal years in the lives of important artists is marked by a profound stylistic change with respect both to their earlier work and to the work of their contemporaries. The aim of our paper is to test this theory by combining two research methods, the qualitative analysis of literary criticism and the quantitative analysis of stylometry.

We chose three representative writers of German literature, Goethe, Musil, and Kafka, and analyzed their late works with the support of established critical interpretations (Trunz 1990, Monti 2000, Kleinwort 2013).

For each author, we ran a double quantitative analysis:

1) an "internal" analysis, where we compared the early, middle, and late works of the same author;

2) an "external" analysis, where we compared the style of the author (early, middle, and late) to reference corpora of contemporary authors and assessed relative deviations from them.

These quantitative analyses were performed through the combination of three di erent methods:

a) Stylometric (Network) Analysis;

b) Stylometric (Zeta) Analysis;

c) Semantic Analysis.

In method (a) we used Cosine Delta distance based on 2,000 most frequent word (Evert et al. 2017) to generate a series of network graphs. In method (b) we used the most over- and under-represented words (Schöch et al. 2018) to evaluate the separation (or overlapping) between the texts in the corpus. In method (c) we used the LIWC dictionary (cf. Tausczik & Pennebaker 2010) to evaluate the semantic areas that dominate a text.

"Late Style" seemed to be con rmed by our studies on semantic areas, but it appeared as unsupported by traditional stylometric methods, thus con rming the results already obtained by (Reeve 2018).

For example, analysis (2a) on Goethe shows no clear separation between Goethe's late works and those of other writers published contemporaneously. Quite surprisingly, also, analysis (2b) shows that Goethe's late texts are more connected to their contemporaries than the previous ones, thus contradicting the supposed isolation of late style.

Analysis (1a) on Musil produces results that even more strikingly challenge the supposed distinctiveness of "Late Style": the network graph is divided into two distinct clusters, and genre, rather than chronology, plays the decisive role.

Analysis (1a) on Kafka shows that the "Early" period—and not the "Late"—distinguished itself most strikingly from the others. This discovery was con rmed by analysis (2a) in which Kafka's four early works were clearly isolated from the rest; in addition, while the "Middle" and "Late" works were not as strongly separated, they remained peripheral in the system. Such a nding was con rmed by analysis (2b), as in all three periods no overlap existed between Kafka's work and work published contemporaneously by others.

Quite di erent were the results of the semantic analysis. As for analysis (1c) in Goethe, (Trunz 1990) identi ed four conceptual symbols as representatives of Goethe's "Late Style": light and totality as symbols of the divine; the eye and the cloud as symbols of humankind's position between niteness and in nity. Based on the examples Trunz provided, we extracted and stemmed a series of words connected to these four semantic areas, and we calculated their frequency in the three sub-sections of Goethe's works. Results con rmed Trunz's intuition. As for analysis (2c), we used the LIWC software, which o ered an opportunity to expand semantic areas to one hundred. We structured our experiment in two phases. First, we isolated the eight LIWC categories in which the comparison between "Late" and "Young/Middle" Goethe showed the highest discrepancies; second, we calculated the frequencies of the same categories in the work of other authors. In the majority of these categories (ve out of eight), Goethe's path diverged diametrically from the approach taken by the authors of other contemporaneously published works.

Similar results were obtained for analysis (1c) in both Musil and Kafka, while analysis (2c) did not produce positive results for Musil.

Such multifaceted results, while not allowing to formulate a de nitive answer to the question whether "Late Style" actually exists, suggest that it could be the result of an author's decision to privilege speci c topics more than it derived from the involuntary use of a certain vocabulary. More in general, they con rm that the synergy of literary hermeneutics and stylometry can open new and innovative perspectives on traditional research questions.

Note: Scripts and corpora used for the analyses are available here: https://github.com/SimoneRebora/LateStyle; works cited: https://www.zotero.org/groups/2261837/latestyledhasa/items

Owe Forum on Facebook: Connecting People, Engaging issues and Promoting Cultural and Ethnic Identity

Josephine Alexander

Keywords: Cultural and Ethnic Identity

Abstract

One of the major challenges that plague migrants all over the world is the longing for "home" and a strong desire to nd meaningful ways of connecting with their cultural and ethnic identity. Nigerians all over the world nd di erent ways of keeping their national and cultural a nity

alive. Prior to Web 2.0 revolution and the exponential growth of social media networking and multi-modal dissemination of information, they use occasions such as Nigeria Independent Day, cultural festivals and ceremonious events such as birthday, wedding and anniversary. The development of new technologies and social media platforms such as Facebook, Twitter, Instagram and Youtube has however rendered such occasions less attractive as new social media networks have digitally created spaces for creative initiatives among individuals and communities. An example of such creative initiative is the Owe Forum on Facebook where every indigene of Kabba, Kogi State of Nigeria irrespective of where they are in the world engage interactively. In this paper, I carry out a nine-month detailed analysis of the postings on the forum through online observation, ethnographic notes and screen data in order to demonstrate the pattern of interaction among the members in relation to connecting people, engaging issues and promoting their ethnic and cultural identity. I demonstrate that networked community such as Owe Forum are potentially fertile digital spaces for data collection and research outputs in indigenous knowledge production, cultural studies, ethnic identity as well as indigenous language documen-tation and maintenance.

Digital Humanities and the Expansion of African Knowledge Societies

Justus Roux

Abstract

An eminent scholar in the eld of African languages (…) recently made the following statement:

"African societies are on the brink of changing from postcolonial societies into global knowledge societies. Digitalization and globalization could enhance their transformation from knowledge-consuming to knowledge-producing societies (…) it would open the way for African indigenous knowledge systems to enjoy recognition in the 'North', not to mention in other parts of the Global South." (Wol , 2018) .

The following questions arise from this view:

1 What are the characteristics of global knowledge societies?

2 Where are African societies currently placed on a continuum of change?

3 What role does language play in producing knowledge?

4 How can African Indigenous Knowledge Systems gain wider recognition in a global context?

Characteristics of knowledge societies

The characteristics of a typical knowledge society based on a report of the United Nations (2005) entitled Understanding Knowledge Societies will be discussed with speci c reference to the notion and practice

of current Information Societies following decisions of the World Summit on the Information Society (WSIS, 2016). The point is made that creating e-communities will not necessarily transform into k-communities (knowledge communities). The need for moving beyond an information delivery function to a preservation of cultural heritage will be stressed in the process of transformation.

State of African societies in moving towards Knowledge Societies

Britz et al. (2006) conducted a study investigating whether Africa was moving towards a knowledge society. They come to the conclusion that "… Africa has still far way to go to become a true knowledge society, but that there is hope to successfully transform Africa into a knowledge society." Britz et al. (2006: i). They however seem to simplify the matter as will be discussed in more detail. In 2017 a joint report of the African Union Commission and other partners was released assessing the state of a airs in di erent African countries in its transition to a Knowledge Economy. Details of the report related to Botswana, Namibia, Kenya, Ghana, Malawi and South Africa will be discussed; where a common issue comes to the fore: the role of indigenous languages in the transformation process.

Language and the production of knowledge

In 2003, a general meeting of UNESCO adopted a resolution on Promotion and Use of Multilingualism and Universal Access to Cyberspace that, inter alia, called on members to ensure that citizens have access to the internet, and produce content in local languages for survival of these languages in cyberspace. Referring to this resolution of UNESCO, Finquelievich and Bassi (2013) argues that governments need to "(…) assist in the preservation and propagation of local languages, customs and other elements of their cultural heritage using ICTs." The relationship between language and knowledge production will be discussed in some detail, referring to the role of Digital Humanities in creating knowledge in di erent languages.

Challenges related to African Indigenous Knowledge Systems

An awareness of indigenous knowledge in many post-colonial countries has given rise to the study of Indigenous Knowledge Systems which, inter alia, strive for the expression of knowledge in a particular vernacular. It will be argued that while it is absolutely necessary to capture and disseminate particular types of knowledge in local languages, it is equally necessary to open up these knowledges for wider communities, hence contributing towards a trove of global knowledge. In this instance particular language technologies have a pertinent role to play.

The presentation will be concluded with a potential example of generating new knowledge in the domain of African Indigenous Knowledge Systems (AIKS) through digital means.

References

Britz, J.J., Lor, P.J., Coetzee, E.M.I. and Bester, B.C. 2006. Africa as a knowledge society: a reality check. International information and library review, 38: 25-40.

Friederici, N., Ojanperä, S. and Graham, M. 2017. The impact of connectivity in Africa: Grand visions and the mirage of inclusive digital development. The Electronic Journal of Information Systems in Developing Countries (EJISDC) 79, 2, 1-20.

Wol , H. E. 2018. Developing Knowledge Societies: Africa Needs a Linguistic Revolution. Blog. https://historyofknowledge.net/2018/02/21/ (accessed 15.03.2018).

Report

United Nations. 2005. Understanding Knowledge Societies. Department of Economic and Social A airs, Division for Public Administration and Development Management. Sales No. E.04.II.C.1, New York.

Workshops

Semantic Web Approaches for Cultural Heritage Modelling

Franziska Pannach

Keywords: Semantic Web, Ontologies, Knowledge Engineering in the Humanities, Annotations, Comparative Folkloristics

Abstract

Workshop Content

In the rst half of the workshop, we will provide an overview about Semantic Web with a special focus on ontologies. The participants will then be introduced to the ontology description language OWL (Web Ontology Language) to create a foundation of understanding about how data is presented in Semantic Web contexts. The workshop aims to delineate how humanities research can bene t from a combination of methods from knowledge engineering and natural language processing. In the second part of the workshop, the users will be introduced to one speci c example of an ontology that can illustrate how the forementioned connection can be realised. Together, we will discuss possible applications that focus on the participants' areas of interest. Our ontology is designed as a tool for annotating and investigating folktales in various languages. It follows the approach of Russian formalist Vladimir Propp [Propp, 1968].

According to Propp, Russian magic tales are made up of a limited set of 31 functions that can be used to describe a tale. These functions follow a strict sequential order. Propp also claims that all of Afanasyev's [Afanasyev, 1916] tales he studied are of one type in regard to their structure, i. e. their sequence of narrative functions [Propp, 1968]. During the workshop, the participants will be introduced to Propp's theory in more detail. We have applied this theory to a small number of English translation of African tales [Smith, 1989], [Greaves, 1990], [Scheub, 2005]. However, we wish to expand the ontology to tales in native African languages.

Our ontology has been carefully modelled using a description logic foundation that de nes the logical interaction between agents and Proppian functions.

Subsequently, it has been implemented in OWL and will be published as a web application in January 2019. It allows analyses on an abstract plot level and on the level of characters and their interaction. A small set of data properties has been de ned to allow representing the functions, tales and publications within the annotation. We intentionally limited the number of data properties to ensure that the ontology remains a useful annotation tool. Our ontology also models concepts that illustrate the family relationship of character‘s in a story [Koleva et al., 2012]. While these additional concepts and relationships are only based on Propp in the sense that they appear in the functions, they are nevertheless crucial elements for the analysis of the tales. So far, ten Bantu tales in their English translation have been annotated by four di erent annotators for test purposes. Although only one of the annotators had previous annotating experience, little assistance was needed for populating the ontology, which is an indicator for the usability of the system. The participants will use the tool in a hands-on session for annotation and query purposes. In addition, we will demonstrate the use of the free Protégé software [Musen, 2015] with which individualized

ontologies can be easily designed and implemented. We believe this workshop can help us learn from the participants, e.g. in which way they nd tools of this sort useful for their speci c area of interest, which needs they have for applying Semantic Web approaches in their research, and how we can establish collaborations from scholars of Computer Science and Humanities.

Target Group

The workshop is targeted at scholars from the Humanities, especially from the eld of Folkloristics, Literature, Media and Cultural Studies and Lingustics.

Furthermore, ontologies have been used to create interesting applications for Historians, e.g. in the eld of Heraldry [Hiltmann, 2015]. It can also be of interest for researchers and sta from Cultural Heritage institutions such as museums or libraries who want to broaden their knowledge about Digital Humanities, Semantic Web and alternative ways of publishing structural data.

Organizatorical Requirements

The workshop is aimed to cover a four-hours session, with two hours for the theoretical foundations, one hour of practical work on the ontology and an hour to exchange ideas about possible applications for the participants research questions. The workshop can take place in a computer-lab with connection to the internet or in a classroom venue if participants bring their own devices.

An ideal group for the knowledge and idea exchange would be 15 participants.

The session should not have more than 20 participants. Since all the basic ideas of Semantic Web, ontologies and the use of the web application/Protégé will be covered, no prior knowledge of the topic is needed. However, we would like to warmly invite interested scholars who have worked with semantic web applications before to come and share their experiences.

References

[Afanasyev, 1916] Afanasyev, A. N. (1916). Russian Folk-Tales. E. P. Dutton & Company, New York.

[Greaves, 1990] Greaves, N., editor (1990). When the Hippo Was Hairy. Lutterworth Press.

[Hiltmann, 2015] Hiltmann, T. (2015). Digital heraldry: Digitisation and dissemination of the european heraldic heritage.

[Koleva et al., 2012] Koleva, N., Declerck, T., and Krieger, H.-U. (2012). An ontology-based iterative text processing strategy for detecting and recognizing characters in folktales. In Meister, J. C., editor, Digital Humanities 2012 Conference Abstracts, pages 467–470. University of Hamburg, Hamburg University Press.

[Musen, 2015] Musen, M. A. (2015). The protégé project: a look back and a look forward. AI Matters, 1(4):4–12.

[Propp, 1968] Propp, V. (1968). Morphology of the Folktale, volume 10. University of Texas Press.

[Scheub, 2005] Scheub, H., editor (2005). African Tales. Univ. of Wisconsin Press.

[Smith, 1989] Smith, A. M., editor (1989). Children of Wax: African Folk Tales. Canongate.

Designing inclusive and accessible platforms, interfaces and content for persons with low vision

Gustaf Tempelho

Keywords: Accessibility, Assistive Technology, Low vision, Universal Design, Interface Design

Abstract

Our interactions with computing devices rely heavily on our visual capability. Because of this we often fail to make provision for users with more speci c needs. This becomes a concern when we develop or design new digital platforms, applications, websites and content. In this workshop we take look at the basic principles of Universal Design and accessibility and why it should be prioritised when creating a new platform or publishing content. The aim of this workshop is to create awareness and to provide a basic grid for the evaluation of accessibility in digital environments with the focus on users with low vision. We will discuss the user experience of digital platforms and interfaces in order to identify the speci c needs and requirements of users with low vision. Based on principals identi ed in the workshop I will demonstrate how assistive technology works (ARIA) and how it could be applied to websites and web applications. We will also discuss how visual content such as images and lm is made accessible to people with low vision.

hours

10 participants

Interest in: Digital projects, interface design and multi-modal content.

Software used: Brackets (freeware), Dream Weaver (Adobe)

Facilities required: Projector, sound and podium/desk. Personal computer for participants (optional)

Social Sciences Data Carpentry at DHASA Conference 2019

Martin Dreyer

Keywords: Data Carpentry, Social Sciences, Python, OpenRe ne

Abstract

Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research. Its target audience is researchers who have little to no prior computational experience, and its lessons are domain speci c, building on learners' existing knowledge to enable them to quickly apply skills learned to their own research after completing the workshop. The lesson material as well as software used during the workshops are available to anyone who would like to access it afterward. Participants will be encouraged to help one another and to apply what they have learned to their own research problems. The workshops are aimed at graduate students and researchers and need no prior knowledge in the subject matter or software.

For more information on what we teach and why, please see our paper “Good Enough Practices for Scienti c Computing”.

The workshop will take place over a two day period, during this time participants will be taught to clean their data using spreadsheets (Data Organisation using Spreadsheets), and do mass manipulation on data using OpenRe ne (OpenRe ne for Social Sciences). The second day of the workshop will teach the participant to analyse data in Python using the Data Analysis and Visualization with Python for Social Scientists lesson material.

Venue requirements for the workshop is a classroom setup with internet access to all participants. Participants are required to bring their own laptops along as the software needed will be installed on each individuals laptop and this is all done during the workshop. As stated previously participants need not have any prior knowledge on the software tools or techniques taught in the workshop.

Virtual Presentations

Virtual presentation

Nadezhda Povroznik

Keywords: Sustainability, digital project, information system, cultural heritage, education

Abstract

Digital humanities projects are often costly in terms of development, building and maintenance. Upon completion of funding, many projects are not continued, despite their demonstrated value for research. The authors of the project "Sustainability in DH: Collaboration and Community" o er arguments in favour of crowdsourcing for the development and maintenance of project activities [1]. Many others have engaged with these research questions over the past decade, but the question of sustainability suggests that new approaches are needed.

The experience of the Center for Digital Humanities at Perm University, Russia, shows that sustainable development of digital projects is possible even after the completion of the nancing of the main stage of the project, and that moreover, it is possible the further develop projects without risking sustainability.

Digital projects aimed at creating information resources and enabling various kinds of research analysis may be able to demonstrate their value in the early phases of a research project, but then have di culty in ensuring continuity unless this has been planned from the outset. In 2013-2015 the Center for Digital Humanities at Perm University, under support of the RFBR, implemented a project aimed at studying history-oriented information systems. A digital catalog was created, which at that time included more than 800 such systems from around the world [2]. The value and relevance of the catalog was apparent from the start. The structure of the description of historical information resources was expanded and the metadata related to the cultural heritage, including the resources of the GLAM sector, were detailed [3].

However, in terms of ensuring sustainability, the question of further expanding and utilising the system involved development of the DigitalHustory.ru platform and its involvement in education. Within a number of disciplines supported by the Center, there are areas related to digital heritage, including the analysis of information resources of digital cultural heritage and their use in various elds. It is important to emphasize the importance of the work of students as future professionals in the humanities with real systems, and not with abstract forms or only prototypes. Practical classes with students and teachers at various levels began to be conducted using the administrative environment of the DigitalHistory.ru platform.

Thus, the experience of the Center shows that the progressive and sustainable development of digital humanities projects is possible after the completion of funding, especially through linking with educational objectives. Continuing the project life cycle, including its extension, improvement and development, depend largely on the creation of a community around such a project, a professional core represented by developers and various users of the system. These and other aspects of sustainability of

digital humanities projects will be discussed in this paper.

References

[1] Project “Sustainability in DH: Collaboration and Community”. URL: https://digitalhumanities.berkeley.edu/blog/15/05/01/project-sustainability-dh-collaboration-and-community

[2] History-oriented information systems. DigitalHistory. URL: http://digitalhistory.ru

[3] Povroznik N. (2018) Virtual Museums and Cultural Heritage: Challenges and Solutions // Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference Helsinki, Finland, March 7-9, 2018. URL: http://ceur-ws.org/Vol-2084/short14.pdf

Submission 127

Martin Dreyer

Keywords: Data Carpentry, Social Sciences, Python, OpenRe ne

Abstract

Data Carpentry develops and teaches workshops on the fundamental data skills needed to conduct research. Its target audience is researchers who have little to no prior computational experience, and its lessons are domain speci c, building on learners' existing knowledge to enable them to quickly apply skills learned to their own research after completing the workshop. The lesson material as well as software used during the workshops are available to anyone who would like to access it afterward. Participants will be encouraged to help one another and to apply what they have learned to their own research problems. The workshops are aimed at graduate students and researchers and need no prior knowledge in the subject matter or software.

For more information on what we teach and why, please see our paper “Good Enough Practices for Scienti c Computing”.

The workshop will take place over a two day period, during this time participants will be taught to clean their data using spreadsheets (Data Organisation using Spreadsheets), and do mass manipulation on data using OpenRe ne (OpenRe ne for Social Sciences). The second day of the workshop will teach the participant to analyse data in Python using the Data Analysis and Visualization with Python for Social Scientists lesson material.

Venue requirements for the workshop is a classroom setup with internet access to all participants. Participants are required to bring their own laptops along as the software needed will be installed on each individuals laptop and this is all done during the workshop. As stated previously participants need not have any prior knowledge on the software tools or techniques taught in the workshop.

Poster Presentations

Spaces, places and collaboration: Transcending the lecture venue

Liezel Blomerus

Keywords: Academic learning support, engagement, enhancement, extension. Learning communities, Personal Learning Environment

Abstract

Academic learning support embedded in a social constructivist learning community approach could foster a culture of learning through the establishment, utilisation in integration of di erent learning spaces. Literature does however suggest that the sharing of practice through institutionally provided platforms are failing to link knowledge assets to people, communities and di erent modes of learning as well as support the development of social networks for learning. The above highlighted the need for an altered approach to the structuring and implementation of academic support in di erent learning spaces, especially if one considers the failures of learning management systems and the availability of free and de facto resources and natural settings that allow for pedagogical a ordances. According to numerous writers, the concept of Personal Learning Environments (PLEs) o ers great potential as a pedagogical approach in answer to the previous.

Against the above background, the Triple E framework of Keren-Kolb (2013) was tapped into to delineate an instructional strategy best suited to t this context of Anthropology modules presented in an Access Programme, at the University of the Free State. This framework allow speci c attention to respectively engagement, enhancement and extension. Although this framework aim to enable technology enhanced learning, its overarching contribution was not intended to be limited to technology enhanced learning; as well as an exclusive dependency on the technological dimension. The concepts "engagement", "enhancement" and "extension" did however lend a useful and appropriate lens to conceptualise academic learning support.

Engagement entail more than just the involvement and participation of students through collaborative and cooperative methods. It involves and accent behavioural, emotional and cognitive components and processes (cf. Trowler, 2010, p. 5). Enhancement stresses the importance of support and the sca olding of content. Extension require the deliberate creation and use of various learning experiences in various and multiple learning spaces (cf. Keren-Kolb, 2013).

The purpose of this poster presentation is to visually portray the process and possibilities of outcomes in the testing of the Triple E framework as instructional strategy, deliberately chosen to integrate the PLE in the formal academic learning support environment and to further contribute, foster and enhance the PLE. Based on an action based research approach this paper suggest the Triple E framework to o er great potential to utilize and integrate di erent reimagined institutionally structured learning spaces, to contribute to engagement that foster behavioural, emotional and cognitive processes and to cater for contextual personal development and empowerment of a student.